ON MEASURES OF CONTINGENCY

By

Frank M. Weida

1. *Introduction.* When we deal with the problem of relationship of attributes, we may classify each attribute into a number of groups. To illustrate: If the attributes are x_i $(i=1,2,3,\cdots,n)$ and if the group belonging to X_1 is x_1^i $(i=1,2,3,\cdots,m_1)$, that belonging to X_2 is x_2^j $(j=1,2,3,\cdots,m_2)$, ..., that belonging to X_i is x_i^k $(k=1,2,3,\cdots,m_i)$, ..., we may form an $m_1\times m_2\times\cdots\times m_i\times\cdots$ table which contains $m_1\times m_2\times\cdots\times m_i\times\cdots$ compartments. In this fashion, it is possible to distribute the total frequency of the "universe" or the "sub-universe" into sub-groups which correspond to these $m_1\times m_2\times\cdots\times m_i\times\cdots$ compartments.

For such situations, Pearson[1] and others[2] have suggested certain measures of relation between the attributes. We shall in this paper be interested primarily in Pearson's measures of contingency. In the case of two attributes, Pearson proceeds as follows: Suppose that A is any attribute and let it be classified into the groups A_i $(i=1,2,3,\cdots,s)$ and let B be another attribute classified into the groups B_j $(j=1,2,3,\cdots,t)$. Let the total number of individuals examined be N. Now, the probability a-priori of an individual falling into the respective groups A_i is n_i/N where n_i is the number which fall into A_i. Again, if m_j is the number which fall into B_j, then the probability a-priori of an individual falling into the respective groups B_j is m_j/N where m_j is the number which fall into B_j. If the attributes are independent in the probability sense, then, if N pairs of attri-

[1] Pearson, Karl, "On the Theory of Contingency and its Relation to Association and Normal Correlation," Drapers' Company Research Memoirs, Biometric Series i.; Dulau & Co., London, 1904.

[2] Yule, G. Udny, "An Introduction to the Theory of Statistics," Charles Griffin & Company, Limited, London, 1927, pp. 17–74.

butes are examined, the number expected in the (ij) compartment is

$$N \cdot \frac{n_i}{N} \cdot \frac{m_j}{N} = \frac{n_i m_j}{N} = \nu_{ij}.$$

Suppose the number observed is n_{ij}. Then, if we allow for the errors of random sampling, $(n_{ij} - \nu_{ij})$ is the departure from independent probability of the occurrence of the groups A, B. Then, any measure of the total departure from independent probability is termed by Pearson a measure of contingency. Consequently, the measure of contingency is some function of the $(n_{ij} - \nu_{ij})$ quantities for the whole table.

Again, for a given

$$\chi^2 = \sum \left\{ \frac{(n_{ij} - \nu_{ij})^2}{\nu_{ij}} \right\}$$

Pearson has shown how to obtain the probability[3] P as a measure to determine how far the observed system is not compatible with a basis of independent probability. He calls $(1-P)$ the *contingency grade* and

$$\phi^2 = \frac{\chi^2}{N}$$

the *mean square contingency*. Also,

$$\psi = \frac{\Sigma (n_{ij} - \nu_{ij})}{N}$$

is the *mean contingency* when Σ refers to summation for all positive terms.

In his theory of contingency, Pearson appears to use the definition of probability used in practically all treatises on the subject.

[3] Pearson, Karl, "On the criterion that a given system of deviations from the probable in the case of correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," Phil. Mag, Series V. 1. 157–175.

This definition excludes the whole field of statistical probability. It appears fairly obvious that the development of statistical concepts is approached more naturally from a limit definition for probability than from the familiar definitions suggested by games of chance. It is the purpose of this paper to improve the treatment of Pearson's theory of contingency and make it more elegant for theoretical as well as empirical discussions. To accomplish this we make use of the notion of *characteristic function*[4] and a definition of probability that includes all forms of probability. It is believed that we have thus idealized Pearson's conception of contingency. We discuss multiple as well as partial contingency. We also consider briefly the case of certain dependent events and the concept of mutual exclusiveness, as well as the concept of connection.

2. *Definitions and assumptions.* In our discussion we need and use the following definitions and assumptions:[5]

Assumption I. If an event which can happen in two different ways be repeated a great number of times under the same essential conditions, the ratio of the number of times that it happens in one way to the total number of trials, will approach a definite limit as the latter number increases indefinitely.

Definition I. The limit described in assumption I we call the *probability* that the event shall happen in the first way under these conditions.

Assumption II. If an event can happen in a certain number of ways, all of which are equally likely, and if a certain number of these be called favorable, then the ratio of the number of favorable ways to the total number is equal to the probability that the event will turn out favorably.

Assumption III. If an event depend on n independent varia-

[4] The characteristic function of A is that function which is equal to unity for the elements of A and zero elsewhere. Usually A is assumed to be a sub-class of some class on which the characteristic function is defined.

[5] Coolidge, J. L., "An Introduction to Mathematical Probability," The Clarendon Press, 1925, pp. 1–12.

bles $X_1, X_2, \cdots\cdots, X_n$ which can vary continuously in an n dimensional continuous manifold, there exists such an analytic function $F(X_1, \cdots\cdots, X_n)$ that the probability for a result corresponding to a group of values in the infinitesimal region

$$X_1 \pm \tfrac{1}{2} dX_1 \;,\quad X_2 \pm \tfrac{1}{2} d\, X_2 , \cdots\cdots , \quad X_n \pm \tfrac{1}{2} d\, X_n .$$

differs by an infinitesimal of higher order from

$$F(X_1, X_2, \cdots, X_n)\, d\, X_1\, d\, X_2 \cdots d\, X_n .$$

Definition II. If a variable X take the different values X_i $(i = 1, 2, \cdots, n)$ with the respective probabilities p_i $(i = 1, 2, \cdots, n)$ and these are all the possible values for that variable, then

$$\sum_{i=1}^{n} p_i X_i$$

is called the *mean value* of the variable X.

Definition III. Two variables are said to be independent if the probability that one lie close to a given value is independent of the value of the other.

3. *Pearson's mean square contingency.* Let the attributes be X and Y. Let ϕ_{ij} be the number of individuals having the group value X_j of X and Y_i of Y. The total number of individuals having the group value Y_i of Y is ϕ_{ij}^{j} [6] and the total number of individuals having the group value X_j of X is ϕ_{ij}^{i}. The total number of individuals examined then is ϕ_{ij}^{ij}.

Now, suppose it is true that

$$F_{ij} = \phi_{ij}^{i}\, \phi_{ij}^{j} , \tag{1}$$

where $F_{ij} = \phi_{ij}^{ij}\, \phi_{ij}$. Let $\overline{F}_{ij}$, $\overline{\phi}_{ij}^{ij}$, $\overline{\phi}_{ij}$, $\overline{\phi}_{ij}^{i}$, $\overline{\phi}_{ij}^{j}$ be, respectively, the *mean values of* F_{ij}, ϕ_{ij}^{ij}, ϕ_{ij}, ϕ_{ij}^{i}, ϕ_{ij}^{j}.

[6] A repeated index means summation for all possible values of such repeated index.

Since, in the case of independence, the mean of the product is the product of the means,[7] we have

(2) $$\bar{F}_{ij} = \bar{\phi}^{i}_{ij} \; \bar{\phi}^{j}_{ij} .$$

Now, if ϕ_{ij} is the characteristic function of the observation, ϕ_{ij} has the value unity if the event succeeds and zero if the event fails. Let p_{ij} be the probability that the event succeeds and q_{ij} the probability that the event fails. Then, the mean value $\bar{\phi}_{ij}$ of ϕ_{ij} is given by

(3) $$\bar{\phi}_{ij} = p_{ij} \cdot 1 + q_{ij} \cdot 0 = p_{ij} .$$

Similarly,

(4) $$\bar{\phi}^{i}_{ij} = p^{i}_{ij} \cdot 1 + q^{i}_{ij} \cdot 0 = p^{i}_{ij} .$$

(5) $$\bar{\phi}^{j}_{ij} = p^{j}_{ij} \cdot 1 + q^{j}_{ij} \cdot 0 = p^{j}_{ij} .$$

(6) $$\bar{\phi}^{ij}_{ij} = p^{ij}_{ij} \cdot 1 + q^{ij}_{ij} \cdot 0 = p^{ij}_{ij} .$$

But $p^{ij}_{ij} = 1$, hence, in the case of independence, $\bar{F}_{ij} = p_{ij}$. Hence, from (2), (3), (4), (5), and (6), in the case of independence, we have

(7) $$p_{ij} = p^{i}_{ij} \cdot p^{j}_{ij} .$$

In the case of dependence, we have that

(8) $$p_{ij} = M(\phi^{i}_{ij} \; \phi^{j}_{ij}) \neq p^{i}_{ij} \; p^{j}_{ij} ,$$

where $M(\phi^{i}_{ij} \; \phi^{j}_{ij})$ is the mean value of $\phi^{i}_{ij} \; \phi^{j}_{ij}$ [8].

The quantity $(p_{ij} - p^{i}_{ij} \; p^{j}_{ij})$ represents the departure between the mean value ϕ_{ij} has and that which it should have in the case of independence.

Let us now consider the square of the departure relative to

[7] Coolidge, J. L., "An Introduction to Mathematical Probability," The Clarendon Press, 1925, p. 62.

[8] Tschuprow, A. A., "Grundbegriffe und grundprobleme der Korrelationstheorie," B. G. Teubner, Berlin, 1925, pp. 39–63.

$p_{ij}^{i}\ p_{ij}^{j}$, namely,

$$\psi_{ij}^{2} = \frac{(p_{ij} - p_{ij}^{i}\, p_{ij}^{j})^{2}}{p_{ij}^{i}\, p_{ij}^{j}}.$$

For all cases, we have

$$\Phi^{2} = (\psi_{ij}^{2})^{ij}, \tag{9}$$

which is Pearson's *mean square contingency* and $\phi_{ij}^{ij}\,\Phi^{2} = \chi^{2}$.

Hence, it appears that we may interpret Pearson's mean square contingency as a coefficient of dispersion, namely, a measure of the deviation between the mean or expected number a cell should have in the case of independence and the mean or expected number it actually has relative to the mean or expected number a cell should have in the case of independence as a unit of measure summed for all cells.

4. *Multiple and partial contingency.* In the case of three variables, suppose that it is true that

$$F_{ijk} = \phi_{ijk}^{i} \cdot \phi_{ijk}^{j} \cdot \phi_{ijk}^{k}, \tag{10}$$

where $F_{ijk} = \phi_{ijk}^{ijk}\ \phi_{ijk}$.

As before, in the case of independence,

$$\bar{F}_{ijk} = \bar{\phi}_{ijk}^{i} \cdot \bar{\phi}_{ijk}^{j} \cdot \bar{\phi}_{ijk}^{k}. \tag{11}$$

Again, if ϕ_{ijk} is the characteristic function of the observation,

$$\bar{\phi}_{ijk} = p_{ijk};\ \bar{\phi}_{ijk}^{i} = p_{ijk}^{i};\ \bar{\phi}_{ijk}^{j} = p_{ijk}^{j};\ \bar{\phi}_{ijk}^{k} = p_{ijk}^{k};\ \bar{\phi}_{ijk}^{ijk} = p_{ijk}^{ijk} = 1. \tag{12}$$

From (10), (11), and (12), in the case of independence, we find that

$$(13) \qquad p_{ijk} = p^{i}_{ijk} \cdot p^{j}_{ijk} \cdot p^{k}_{ijk},$$

and in the case of dependence, we have

$$(14) \qquad p_{ijk} = M(\phi^{i}_{ijk} \cdot \phi^{j}_{ijk} \cdot \phi^{k}_{ijk}) \neq p^{i}_{ijk}\, p^{j}_{ijk}\, p^{k}_{ijk}.$$

The quantity $(p_{ijk} - p^{i}_{ijk} \cdot p^{j}_{ijk} \cdot p^{k}_{ijk})$ represents the departure between the mean value ϕ_{ijk} has and that which it should have in the case of independence.

We now consider the square of the departure relative to $p^{i}_{ijk} \cdot p^{j}_{ijk} \cdot p^{k}_{ijk}$, namely,

$$\psi^{2}_{ijk} = \frac{(p_{ijk} - p^{i}_{ijk} \cdot p^{j}_{ijk} \cdot p^{k}_{ijk})^{2}}{p^{i}_{ijk} \cdot p^{j}_{ijk} \cdot p^{k}_{ijk}}.$$

For all cases, we have

$$(15) \qquad \Phi^{2} = (\psi^{2}_{ijk})^{ijk},$$

which we call the ***mean square multiple contingency*** in the case of three variables or attributes.

In general, in case we have n attributes:

$$(16) \qquad \psi^{2}_{i_1 i_2 \cdots i_n} = \frac{(p_{i_1 i_2 \cdots i_n} - p^{i_1}_{i_1 i_2 \cdots i_n} - \cdots - p^{i_n}_{i_1 i_2 \cdots i_n})^{2}}{p^{i_1}_{i_1 i_2 \cdots i_n} \cdots\cdots p^{i_n}_{i_1 i_2 \cdots i_n}},$$

and for all cases:

$$(17) \qquad \Phi^{2} = (\psi^{2}_{i_1 i_2 \cdots i_n})^{i_1 i_2 \cdots i_n},$$

which we call the ***mean square multiple contingency*** in the case of n attributes.

Let us again consider the case of three attributes. We may write

$$\Phi^2 = \left\{ \left(\psi_{ij}^2 \right)_k^{ij} \right\}^k = \left\{ \left(\psi_{ik}^2 \right)_j^{ik} \right\}^j = \left\{ \left(\psi_{jk}^2 \right)_i^{jk} \right\}^i .$$

For a given k ,

$$\Phi_k^2 = \left(\psi_{ij}^2 \right)_k^{ij} \tag{18}$$

is the *partial mean square contingency* between two attributes for an assigned third attribute.

If $\Phi_k = 0$ for every k $(k = 1, 2, 3 \cdots)$, then

$$\Phi^2 = \left(\Phi_k^2 \right)^k = 0.$$

Similarly, if Φ_i^2 and Φ_j^2 are zero for every i and every j , respectively, then

$$\Phi^2 = \left(\Phi_i^2 \right)^i = 0 \quad \text{, and}$$

$$\Phi^2 = \left(\Phi_j^2 \right)^j = 0$$

We have thus proved the theorem, namely,

Theorem 1: The necessary and sufficient condition for the three attributes to be independent is that

$$\begin{cases} \Phi^2 = \left(\Phi_k^2 \right)^k = 0 , & \text{or} \\ \Phi^2 = \left(\Phi_i^2 \right)^i = 0 , & \text{or} \\ \Phi^2 = \left(\Phi_j^2 \right)^j = 0 . & \end{cases} \tag{19}$$

It is fairly easy to see that in the case of n attributes, we have

$$\Phi^2 = \left\{ \left(\psi_{i_1 i_2} \right)_{i_3 \cdots i_n}^{i_1 i_2} \right\}^{i_3 \cdots i_n}$$

For a given set $i_3, i_4, \cdots, i_n$

$$\Phi^2_{i_3 i_4 \cdots i_n} = (\Psi^2_{i_1 i_2})^{i_1 i_2}_{i_3 i_4 \cdots i_n} \tag{20}$$

where $\Phi_{i_3 i_4 \cdots i_n}$ is the *partial mean square contingency* between two attributes for an assigned set of $(n-2)$ attributes.

If $\Phi_{i_3 \cdots i_4} = 0$ for any pair i_1, i_2, and for every associated set $i_3, i_4, \cdots\cdots, i_n$, then

$$\Phi^2 = (\Phi^2_{i_3 i_4 \cdots i_n})^{i_3 \cdots i_n} = 0.$$

Hence, we have the

Theorem 2: The necessary and sufficient condition for complete independence in the case of n attributes is that for every pair i_1, i_2, it is true that

$$\Phi^2 = (\Phi^2_{i_3 i_4 \cdots i_n})^{i_3 i_4 \cdots i_n} = 0. \tag{21}$$

Again, it is fairly easy to see that in general different values assigned to the set $i_3, i_4, \cdots, i_n$ will result in corresponding different values for $\Phi^2_{i_3 i_4 \cdots i_n}$. Hence, if ${}_w\Phi_{i_3 i_4 \cdots\cdots i_n}$ is the weighted arithmetic mean of these different values where the respective weights are the relative numbers of individuals in each sub-set, then we say that

$${}_w\Phi^2_{i_3 i_4 \cdots i_n}.$$

is the *partial mean square measure of contingency.*

5. *Mean square dependence.* Rietz[9] invented games of chance which give a meaning to correlation in pure chance. The writer believes it important at least formally to propose a measure of

[9] Rietz, H. L., "Urn schemata as a basis for the development of correlation theory," Annals of mathematics, Vol. 21, 1919–20, pp. 306–322.

dependence based upon a probability schemata. As before, let the attributes be X and Y.

Let us assume that

$$F_{ij} = F(\phi^i_{ij}, \phi^j_{ij}, i, j). \text{ Then,}$$

$$\bar{F}_{ij} = \bar{F}(\phi^i_{ij}, \phi^j_{ij}, i, j), \text{ whence,}$$

$$p_{ij} = P_{ij},$$

where P_{ij} is the mean value of $F(\phi^i_{ij}, \phi^j_{ij}, i, j)$ and p_{ij} is the mean value of F_{ij}.

The quantity $(p_{ij} - P_{ij})$ represents the departure from dependence for the particular $F(\phi^i_{ij}, \phi^j_{ij}, i, j)$ under discussion. We now form the quantity D_{ij} defined as

$$D^2_{ij} = \frac{(p_{ij} - P_{ij})^2}{P_{ij}}, \tag{22}$$

which is the square of the departure relative to P_{ij}.

For all cases, we have

$$\delta^2 = (D^2_{ij})^{ij}, \tag{23}$$

which we call the *mean square dependence*.

Our concept of dependence may be extended to cases of more than two attributes and measures of multiple as well as partial dependence may be obtained in an analogous fashion. It thus appears that we have, at least formally, a general criterion for dependence and an approach to a general criterion which may serve as a *measure of goodness of fit*.

We also note that in every contingency table the events designated by the p_{ij} or P_{ij} are *mutually exclusive* for every i and j.

6. *A measure of connection.* We here propose to idealize Gini's measure of connection which has been fully discussed by

the writer elsewhere.[10] Gini's measure of connection is of interest and importance since one of his special indices of connection is Pearson's correlation ratio and one of his special indices of concordance is Pearson's correlation coefficient. These facts are established in my paper referred to above.

As before, let ϕ_{ij} represent the number of individuals having the group value X_j of X and Y_i of Y in case we have the two attributes X and Y. The total number of individuals having the group value Y_i of Y is ϕ_{ij}^{j} and the total number of individuals having the group value X_j of X is ϕ_{ij}^{i}. The total number of individuals is ϕ_{ij}^{ij}. The frequencies of Y are distributed according to a set of "partial" groups which correspond to the respective modalities of X. If all the "partial" groups are similar to the "total" group of frequencies of Y, then the distribution of modalities of Y is independent of the modalities of X and Y is not connected with X. In other words, Y is not dependent upon X but is independent of X in the probability sense. Again, if at least one of the "partial" groups is not similar to the "total" group of frequencies of Y, then the distribution of modalities of Y is dependent on the modalities of X and Y is connected with X. In other words, Y is dependent on X and is not independent of X in the probability sense.

We now multiply the frequencies of each "partial" group by a number w_j such that the total frequency of each "partial" group is the same as the number of cases examined. For a given cell, the frequency is then $w_j \phi_{ij}$ and the total frequency of this "partial" group is then $w_j \phi_{ij}^{i} = \phi_{ij}^{ij}$.

Let us now consider the quantity G_{ij} defined by

$$G_{ij} = \phi_{ij}^{j} - w_j \phi_{ij} .$$

The mean value of ϕ_{ij}^{j} is p_{ij}^{j} and the mean value of $w_j \phi_{ij}$ is

[10] Weida, F. M., "On various conceptions of correlation," Annals of Mathematics, Vol. 29, No 3, July 1928, pp. 276–312.

p_{ij} . If M_{ij} is the mean value of G_{ij} then

$$M_{ij} = p_{ij}^{j} - p_{ij} . \tag{24}$$

We now consider a quantity d_j defined by

$$d_j = (|M_{ij}|)^{i}, \tag{25}$$

which is Gini's *simple index of dissimilarity* and may be regarded as the sum of the absolute values of a set of mean values.

We now consider the quantity $\phi_{ij}^{i} d_j$. The mean value of $\phi_{ij}^{i} d_j$ is $p_{ij}^{i} d_j$.

For all cases, the mean value I_{YX} is given by

$$I_{YX} = (p_{ij}^{i} d_j)^{j}, \tag{26}$$

which is Gini's *measure of connection of* Y *on* X . Thus, Gini's measure of connection may be regarded as the mean value of a set of sums of absolute values of mean values. An analagous discussion holds for I_{XY} which is Gini's measure of a connection of X on Y .

It is fairly easy to see that the process may be extended to derive measures of multiple, partial and complete connection. This the writer intends to accomplish at a future date.

7. *Conclusion.* It is believed that we have shown that the theory of contingency, dependence and connection may be based upon a definition of probability that includes all forms of probability. Fluctuations in random sampling appear to be neglected in such a treatment, however the experiments may be carried out with the probability schemata in case we desire the inclusion of fluctuations in random sampling.

The George Washington University